Malays. Appl. Biol. (2015) 44(1): 13–18

SPATIAL VARIATION ASSESSMENT OF MALACCA RIVER WATER QUALITY USING MULTIVARIATE STATISTICAL ANALYSIS

SH. NURSHAFAZILLAH ROSLI[1*], AHMAD ZAHARIN ARIS[1] and NIK M. MAJID[2]

[1]*Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Institute of Tropical Forestry and Forest Product, 43400 UPM Serdang, Selangor, Malaysia*
[*]*Email: qistina_nuur@yahoo.com*

ABSTRACT

This study evaluates the spatial variation of river water quality and identifies major sources of water pollution along Malacca River using cluster analysis, discriminant analysis and principal component analysis. The data sets contain 23 water quality parameters from seven monitoring stations over a ten-year monitoring period (2002-2011). The seven stations were grouped based on similar characteristics of sampling stations using cluster analysis into low-polluted sites, moderately-polluted sites and highly-polluted sites. In discriminant analysis, the original 23 parameters were reduced to 12 and 15 of the most significant pollutants in forward and backward stepwise mode, respectively. In principal component analysis, the results showed that pollution sources for moderately-polluted sites and highly-polluted sites are related to point sources and non-point sources while in low-polluted sites, pollution is mainly due to non-point sources. This study demonstrates the effectiveness of multivariate statistical method for assessment and interpretation of bulky and complex river water quality data in order to design a better supervision network for successful management of water resources.

Key words: Cluster analysis, discriminant analysis, principal component analysis, water quality

INTRODUCTION

The Malacca River flows through the City of Malacca. In recent years, the river has been classified as a polluted river (Class III) according to the Water Quality Index (Department of Irrigation and Drainage, 2013). This problem is due to rapid development along the river that transformed most parts of the forested watershed into urban areas and contributed to both point and non-point sources of pollution (Paul and Meyer, 2011). The point sources are caused by anthropogenic activities such as urbanization, industrial sewage, and wastewater treatment plants (Pejman *et al*., 2009; Nouri *et al*., 2008). Animal husbandry discharge and agricultural activities have also become an issue along the Malacca River. The non-point sources are due to natural processes such as soil erosion and surface runoff (Liao *et al*., 2006; Mahvi *et al*., 2005). The surface runoff carries agricultural, industrial, and domestic discharge directly into the river and raises the degree of river water pollution (Sickman *et al*., 2007; Simeonov *et al*., 2003).

Therefore, monitoring and data analysis statistical programs are required for a reliable assessment of river water quality which is a vital factor affecting ecological systems and human health (Wang *et al*., 2013). However, the degree to which each factor contributes to river pollution is indistinct because the data are bulky and complex (Zhang *et al*., 2009; Singh *et al*., 2004). It is therefore necessary to decrease uncertainty by interpreting spatial variation and identifying the most significant factors of river water pollution (Gazzaz *et al*., 2012). The results could provide meaningful information and help in decision making for improving river water quality (Templ *et al*., 2008; Muxika *et al*., 2007). The aims of the study were to evaluate the similar characteristics of sampling stations by using cluster analysis, to identify the most significant variables for spatial variation by discriminant analysis, and to determine the main pollution sources that affect river water quality of Malacca River based on principal component analysis.

* To whom correspondence should be addressed.

STUDY AREA AND METHODS

Description of study area

The length of Malacca River is about 39.0 km and total catchment area is 608 km^2. In this study, seven water quality monitoring stations were selected (1MN 01, 1MN 04, 1MN 05, 1MN 06, 1MN 12, 1MN 14 and 1MN 17) covering most parts of the Malacca River (Fig. 1).

Water quality parameters

A total of 23 water quality parameters were selected based on the consistency of monitoring by the Department of Environment from 2002 to 2011. The parameters are dissolved oxygen, biological oxygen demand, chemical oxygen demand, pH, temperature, conductivity, suspended sediment, dissolved solid, total solid, turbidity, ammoniacal nitrogen, magnesium, sodium, nitrate, chlorine, phosphate, zinc, calcium, iron, potassium, arsenic, *Escherichia coli* (*E. coli*) and salinity.

Statistical analysis

The selected parameters were analysed using three statistical techniques; cluster analysis discriminant analysis and principal component analysis. Cluster analysis classified the data sets into a group or cluster which has the similarities in water quality characteristics (Gazzaz *et al*., 2012; Juahir *et al*., 2011). Discriminant analysis was then performed on the original data sets using three different modes, mainly standard, forward stepwise and backward stepwise in order to determine the most significant variable of water quality (Shrestha and Kazama, 2007; Kowalkowski *et al*., 2006). The data sets were further analysed using principal component analysis to identify the source of pollution based on physical, chemical, or biological characteristics (Gazzaz *et al*., 2012; Mutihac and Mutihac, 2008).

RESULTS AND DISCUSSION

Site grouping by cluster analysis

In cluster analysis, seven monitoring stations along the Malacca River were grouped into three clusters at $(D_{link}/D_{max}) \times 100 < 200$, represented in a dendrogram (Fig. 2). Four stations (1MN 04, 1MN 05, 1MN 06 and 1MN 17) formed Cluster 1, which represent less polluted sites (LPS) with the highest mean dissolved oxygen values. These stations are located at the upstream of Malacca River and covered the village area namely Durian Tunggal, Pantai Belimbing and Melaka Pindah. These stations also have less development compared to the other sites (Fig. 3), resulting in Class I and Class II Water Quality Index. The pollution was mostly from non-point sources. Cluster 2 (1MN 14) corresponds to a moderately-polluted site (MPS) and located at the midstream with Class II river classification. Cluster 3 (1MN 01 and 1MN 12) is categorized as highly-polluted site (HPS) with the lowest mean dissolved oxygen values. These stations are located

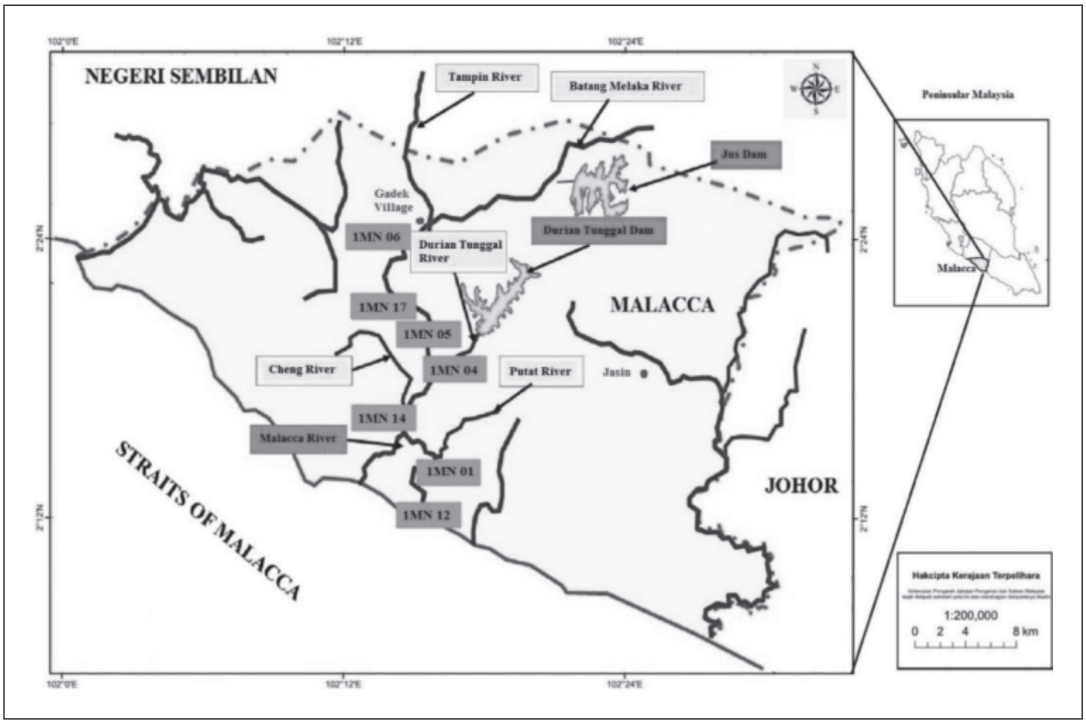

Fig.1. Seven selected stations along Malacca River

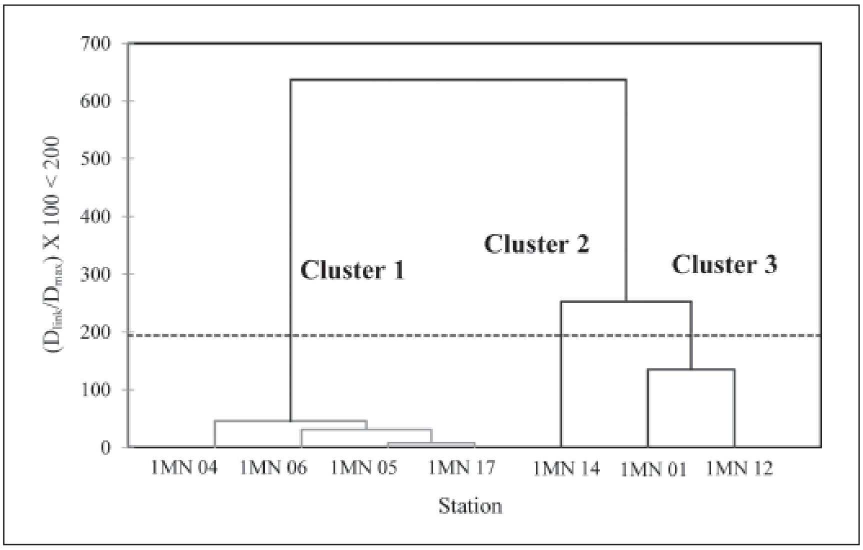

Fig. 2. Dendrogram of Malacca River monitoring stations in three clusters

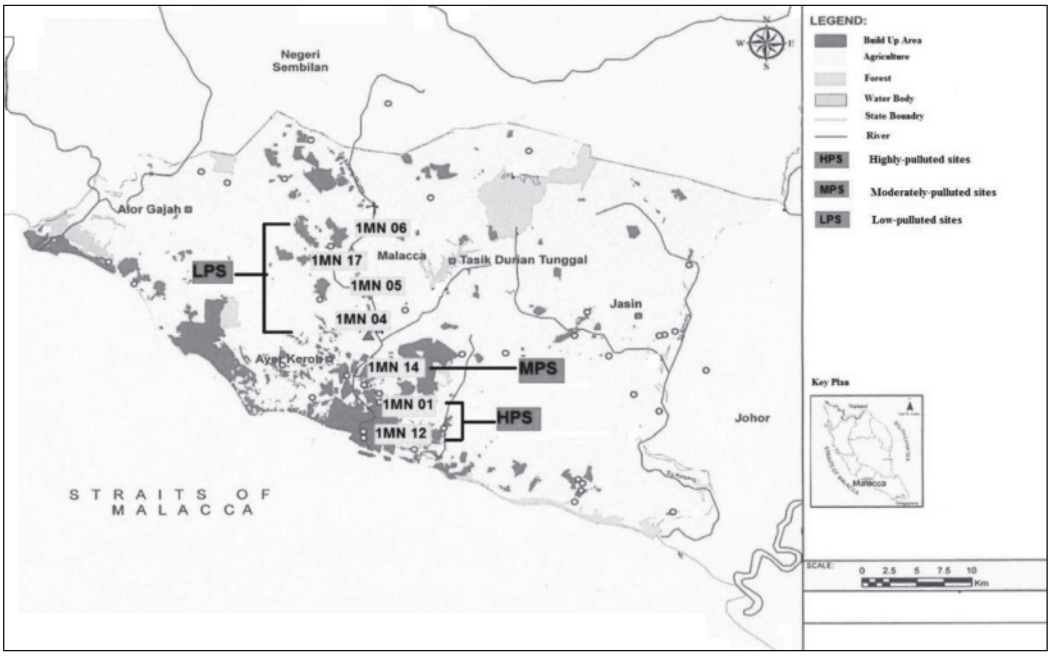

Fig. 3. Classification of monitoring stations at Malacca River
(*Source from Department of Irrigation and Drainage, DID*)

at the downstream with Class III Water Quality Index due to the high population density as well as rapid development activities (Fig. 3). The pollution originated from both point and non-point sources. These stations received the pollution from industrial area such as Batu Berendam, Ayer Keroh and Tasik Utama. The river in Cluster 3 also appeared darker brown and smelly compared to Cluster 1, which was clearer with a light brown colour. In addition, there are tributaries such as Putat River and Cheng River which are polluted and contribute to the pollution of Malacca River.

Spatial variation in water quality

The 23 parameters were further evaluated using discriminant analysis to identify the most significant variables contributing to pollution sources on the river. Based on Table 1, standard mode, forward and backward stepwise mode resulted an accuracy of 94.64%, 94.64% 94.81% indicating a high value of spatial variation, respectively. The initial 23 parameters were reduced to 12 and 15 representative parameters through forward and backward stepwise methods, respectively. Based on the backward stepwise method (Table 1), the spatial analysis

Table 1. Correlation matrix for discriminant analysis of spatial variation

Sampling Sites	Percent of correct (%)	Sites		
		HPS	MPS	LPS
Discriminant analysis standard (23 variables)				
Highly-polluted sites	81.31	87	12	8
Moderately-polluted sites	80.00	4	40	6
Low-polluted sites	99.76	0	1	420
Total	94.64	91	53	434
Forward stepwise (12 variables)				
Highly-polluted sites	81.31	87	13	7
Moderately-polluted sites	80.00	4	40	6
Low-polluted sites	99.76	0	1	420
Total	94.64	91	54	433
Backward stepwise (15 variables)				
Highly-polluted sites	82.24	88	11	8
Moderately-polluted sites	80.00	4	40	6
Low-polluted sites	99.76	0	1	420
Total	94.81	92	52	434

Table 2. Factor loading on significant verifactors (>0.6)

Variables	HPS						MPS						LPS					
	V1	V2	V3	V4	V5	V6	V1	V2	V3	V4	V5	V6	V1	V2	V3	V4	V5	V6
DO											0.8					0.8		
BOD_5				0.7					0.7									
COD									0.7									
SS		0.8						0.9						0.9				
pH	0.7																	
NH_3-NL										0.6								
Temp																		
Turbidity		0.8						0.9						0.9				
DS	0.9						0.9						0.9					
TS	0.9						0.7							0.9				
Mg	0.8												0.8					
Na	0.9						0.9						0.9					
COND	0.9						0.9						0.8					
Salinity	0.9						0.9						0.9					
NO_3																		
Cl^{-1}	0.9						0.7						0.9					
PO_4				0.8						0.6					0.7			
Zn																		
Ca			0.9				0.8											
Fe																		
K			0.9				0.7								0.7			
As																	0.6	
E-coli																		
Cum Var (%)						**76**						**71**						**67**

suggested dissolved oxygen, biological oxygen demand, chemical oxygen demand, pH, temperature, conductivity, salinity, dissolved solid, turbidity, ammoniacal nitrogen, nitrate, phosphate, calcium, iron, and arsenic are the most significant parameters contributing to the three regions; low-polluted sites, moderately-polluted sites and highly-polluted sites.

Sources identification of monitoring sites

Principal component analysis shows strong correlation with pH, DS, TS, Mg, Na, COND, SAL, Cl^{-1}, SS, TUR, Ca, K, PO_4, and BOD_5 at highly-polluted sites (Table 2). High dissolved solid, total solid, suspended solid and turbidity are caused by soil erosion due to land use activities such as sand mining and land clearing along the Malacca River. Salinity shows high salt content in water, resulting from fertilizer runoff. The conductivity shows the presence of ions (Na, Mg and Cl) as mineral components of river water (Wang *et al*., 2013). Calcium and potassium are from chemical components of anthropogenic activities which include domestic and agricultural runoffs. Biological oxygen demand represents the influence of organic pollutants from the use of excessive fertilizer for agriculture purposes. Highly-polluted sites located at the downstream of Malacca River possibly received most of pollution from point and non-point sources. This is due to the rapid urbanization, high population density, and anthropogenic activities. The pollution have originated from wet markets, domestic sewage treatment plant, livestock farms, industrial effluent and surface runoff from agricultural activities.

For moderately-polluted sites, the result shows strong correlation on DS, SS, Na, COND, SAL, Cl^{-1}, Ca, K, SS, TUR, BOD_5, COD, NH_3-NL, PO_4, and DO. Dissolved solid, suspended solid and turbidity factors are due to land clearing, soil erosion of road edges, surface runoff, and agricultural runoff. Biological oxygen demand and chemical oxygen demand may be due to the discharge of municipal and industrial wastes (Juahir *et al*., 2011). Ammoniacal nitrogen and phosphate factors are inorganic nutrients from excessive influx of orchards, agriculture, and livestock wastes. Municipal and industrial sewage are also sources of phosphate because it is an important component in detergents (Vega *et al.*, 1998). Moderately-polluted site is located at the midstream of Malacca River and received pollution mostly from anthropogenic activities including industrial effluent, land clearing, municipal sewage, paint-based industries, and surface runoff from agricultural activities. However, these sources only slightly pollute the Malacca River because of moderately built-up area and less heavy industry in this region compared to highly-polluted sites (Fig. 3).

For low-polluted sites, the results show strong correlation on DS, Mg, Na, COND, SAL, Cl^-, TS, SS, TUR, PO_4 K, DO, and As. High dissolved solid, suspended solid, total solid, and turbidity are due to land use activities in this region which may worsen during the wet season when runoff is carried along the soil and mud into the river. The presence of phosphate and potassium are due to the fertilizer input and farming activities and thus dissolved via surface runoff. High concentrations of dissolved oxygen may be influenced by aeration, photosynthesis, respiration and oxidation from wastes (Zhou *et al*., 2007; Wunderlin *et al*., 2001). Toxic pollutants from pesticides possibly enter the water body via surface runoff, thus contaminating water quality. Therefore, more of the pollution sources in low-polluted sites originate from non-point sources.

CONCLUSION

In this study, three multivariate statistical techniques were successfully employed to evaluate the spatial variation of Malacca River water quality. Cluster analysis grouped the seven monitoring station into three clusters based on similar characteristics of water quality parameter and pollution sources. The initial variables were then reduced to the most significant pollutants by discriminant analysis. Further analysis using principal component analysis provided meaningful information on the major pollutants along Malacca River and their sources. This analysis offer reliable classification of Malacca River water quality and may serve as cost-effective method in water quality assessment. This helps in rapid evaluation of spatial variation of water quality and future monitoring activities.

ACKNOWLEDGEMENTS

We would like to thank the Ministry of Education Malaysia for financial support through the Research University Grant Scheme (91784), Department of Environment (DOE) for the secondary data and to Department of Irrigation and Drainage (DID) for the recommendations and valuable information about Malacca River.

REFERENCES

Department of Irrigation and Drainage, 2013. Annual Report.

Gazzaz, M.N., Yusof, M.K., Ramli, M.F., Aris, A.A. & Juahir, H. 2012. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Marine Pollution Bulletin,* **64(12)**: 688-698.

Juahir, H., Sharifuddin, M.Z. & Kamil, Y. 2011. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring Assessment*, **173**: 625-641.

Kowalkowski, T., Zbytniewski, R., Szpejna, J. & Buszewski, B. 2006. Application of chemometrics in river classification. *Water Research,* **40(4)**: 744-752.

Liao, S., Lai, W., Chen, J., Sheu, J. & Lee, C. 2006. Water quality during development and apportionment of pollution from rivers in Tapeng Lagoon, Taiwan. *Environmental Monitoring and Assessment,* **122(1-3)**: 81-100.

Mahvi, A.H., Nouri, J., Babaei, A.A. & Nabizadeh, R. 2005. Agricultural activities impact on groundwater nitrate pollution. *International Journal of Environmental Science and Technology*, **2(1)**: 41-47.

Mutihac, L. & Mutihac, R. 2008. Mining in chemometrics. *Analytica Chimica Acta*, **612(1)**: 1-18.

Muxika, I., Biorja, A. & Bald, J. 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. *Marine Pollution Bulletin*, **49**: 809-818.

Nouri, J., Karbassi, A.R. & Mirkia, S. 2008. Environmental management of coastal regions in the Capsian Sea. *International Journal of Environmental Science and Technology*, **5(1)**: 43-52.

Paul, M.J. & Meyer, J.L. 2011. Streams in the urban landscape. *Annual Revision Ecological System*, **32(1)**: 333-365.

Pejman, A.H., Nabi Bidhendi, G.R., Karbassi, A.R., Mehrdadi, N. & Bidhendi, E. 2009. Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *International Journal of Environmental Science and Technology,* **6(3)**: 467-476.

Shrestha, S. & Kazama, F. 2007. Assessment of surface water quality using multivariate statistical techniques. A case study of the Fuji river basin, Japan. *Environmental Modeling Software*, **22(4)**: 464-475.

Sickman, J.O., Zanoli, M.J. & Mann, H.L. 2007. Effects of urbanization on organic carbon loads in the Sacramento River, California. *Water Resources Research*, **43**: 10-15.

Singh, K.P., Malik, A., Mohan, D. & Sinha, S. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water Research*, **38**: 3980-3992.

Simeonov, V., Straits, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, Th. 2003. Assessment of surface water quality in Northern Greece. *Water Research*, **37(3)**: 4119-4124.

Templ, M., Filzmoser, P. & Reimann, C. 2008. Cluster analysis applied to regional geochemical data, problems and possibilities. *Applied Geochemistry*, **23(8)**: 2198-2213.

Vega, M., Pardo, R., Barrado, E. & Deban, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, **32**: 3581-3592.

Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., Laura, F.M. & Li, B.L. 2013. Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-environment Research*, **7(13)**: 30-40.

Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Hued, A.C. & Biston, M.A. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba-Argentina). *Water Research*, **35(12)**: 2881-2894.

Zhang, Q., Li, Z.W., Zheng, G.M., Li, J.B., Fang, Y., Yuan, Q.S., Wang, Y.M. & Ye, F.Y. 2009. Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: A case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment*, **152**: 105-121.

Zhou, F., Guo, H., Liu, Y. & Jiang, Y. 2007. Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong. *Marine Pollution Bulletin*, **54**: 745-756.